

Convergence of the D-iteration algorithm: convergence rate and asynchronous distributed scheme

Dohy Hong
Alcatel-Lucent Bell Labs
Route de Villejust
91620 Nozay, France

dohy.hong@alcatel-lucent.com

Fabien Mathieu
INRIA
avenue d'Italie
75014 Paris, France

fabien.mathieu@inria.fr

Gérard Burnside
Alcatel-Lucent Bell Labs
Route de Villejust
91620 Nozay, France

gerard.burnside@alcatel-lucent.com

ABSTRACT

In this paper, we define the general framework to describe the diffusion operators associated to a positive matrix. We define the equations associated to diffusion operators and present some general properties of their state vectors. We show how this can be applied to prove and improve the convergence of a fixed point problem associated to the matrix iteration scheme, including for distributed computation framework. The approach can be understood as a decomposition of the matrix-vector product operation in elementary operations at the vector entry level.

Categories and Subject Descriptors

G.1.3 [Mathematics of Computing]: Numerical Analysis—*Numerical Linear Algebra*; G.1.0 [Mathematics of Computing]: Numerical Analysis—*Parallel algorithms*

General Terms

Algorithms, Performance

Keywords

Linear algebra, numerical computation, iteration, fixed point, eigenvector, distributed computation

1. INTRODUCTION

Today, we are living at the heart of the information age and we are facing more and more data for which it becomes difficult to process using, say, traditional data processing strategies. In such a context, iterative methods to solve large sparse linear systems have been gaining interests in many different research areas and a large number of solutions/approaches have been studied. For instance, iterative techniques gained a significant efficiency by exploiting the sparsity of the information structure (decomposition, partition strategies etc). However, if a near to optimal specific and direct solution can be built for a given problem, it is obviously hard to have one solution that remains optimal for a large classe of problems. One of the most noticeable *generic* improvement was brought by iterative methods combining the Krylov subspace and preconditioning approaches [16], [9], [3].

*The work presented in this paper has been partially carried out at LINCS (www.lincs.fr).

In this paper, we propose a new iterative algorithm based on a simple and intuitive fundamental understanding of the linear equations as diffusions: we believe that this approach may bring significant improvement in a large classe of linear problems. More precisely, we study the fixed point convergence problem in linear algebra exploiting the idea of fluid diffusion associated to the D-iteration [12]. This approach has been initially proposed to solve the PageRank equation [2], [13]. The D-iteration solves $X \in \mathbb{R}^N$ of the equation:

$$X = \mathbf{P}X + B$$

where \mathbf{P} is a non-negative matrix. This includes in particular the case where \mathbf{P} is of spectral radius unity (and $B = 0$).

Solving a linear problem is a very well known theoretical problem and there are a large number of methods that have been proposed, studied and explored. For the general description of existing and/or alternative iteration methods, one may refer to [7], [16], [18], [8], [17]. In particular, for the power iteration method (solving $\mathbf{P}X = X$), the theory is very well known, for instance when \mathbf{P} is associated to an irreducible transition matrix of a Markov chain, X would be its unique stationary distribution (cf. [6], [4]). The convergence of classical iterative schemes, such as Jacobi or Gauss-Seidel or successive over-relaxation method, is also a very well known problem. One usual convergence condition is that (condition expressed for the equation $\mathbf{A}X = B$, \mathbf{A} may be chosen equal to $\mathbf{I} - \mathbf{P}$) \mathbf{A} is strictly diagonal dominant. With other approaches such as Krylov, conjugate gradient and variant methods exploiting the idea of projection and residual minimization, better convergence can be obtained under more restrictive conditions on \mathbf{A} (such as symmetric and positive definite). If the power iteration (or the usual matrix-vector product) can be associated to linear operations on the rows of the matrix (assuming an iteration is defined by the product matrix-vector), the diffusion approach consists in exploiting the columns of \mathbf{P} . The diffusion approach requires in particular that we have to define two state vectors F (fluid) and H (history). In this paper, we revisit the convergence condition of the D-iteration and we show that its convergence can be obtained under a very wide condition on \mathbf{P} , with an upper bound on its convergence rate when the spectral radius of \mathbf{P} is strictly less than unity. One of the main advantage of the diffusion approach is that its distance to the limit is explicitly known and that the convergence speed gain can be intuitively and simply analyzed.

Concerning iterative methods on distributed computation framework, the usual approach is to define a synchronization phase: for instance, in GPU computation [14] or MapReduce framework [5], the synchronization phase is an explicit step of the architecture. For those approaches, the convergence condition is generally the same as for the sequential computation. For an asynchronous distributed computation framework, the convergence may be harder to prove or may simply fail: [15] is one of the first paper proving the convergence of the asynchronous distributed computation of power iterations under pretty wide conditions.

The diffusion approach is well suited to an asynchronous distributed computation (cf. [11, 10]), but its convergence had not been proved formally. In this paper, we formally prove the convergence of the diffusion approach (D-iteration) both for the sequential computation (single processor) and for the distributed computation cases.

We compare the performance of the distributed computations based on the theoretical simulation framework introduced in [15] and show that for sparse large matrices, the case for which the diffusion approach was initially designed, the computation gain of the proposed method compared to the row-based asynchronous parallel computation is significantly large, close to the theoretical optimal efficiency for large sparse matrices.

In Section 2, we define the notations and some general properties. Section 3 gives the formal framework of the D-iteration method and some general results, in particular, the theoretical proofs of the convergence for the sequential approach (single processor). In Section 4, we show that for the most natural three variants of the D-iteration, a simple upper bound of the convergence speed can be obtained when \mathbf{P} can be reduced in a simple multiplicative form. Section 5 gives the proof of the convergence of the distributed scheme. Finally, Section 6 reports and discusses some experimental results.

2. GENERAL FRAMEWORK

We will use the following notations:

- $\mathbf{P} \in (\mathbb{R}^+)^{N \times N}$ a non-negative matrix;
- $\mathbf{I} \in (\mathbb{R}^+)^{N \times N}$ the identity matrix;
- \mathbf{J}_i the matrix with all entries equal to zero except for the i -th diagonal term: $(\mathbf{J}_i)_{ii} = 1$;
- $\Omega = \{1, \dots, N\}$;
- $\mathcal{I} = \{i_1, i_2, \dots, i_n, \dots\}$ a sequence of nodes: $i_k \in \Omega$;
- $\sigma_v : \mathbb{R}^N \rightarrow \mathbb{R}$ defined by $\sigma_v(X) = \sum_{i=1}^N v_i x_i$; if V has no zero entry, we define the norm $|X|_v = \sum_{i=1}^N |v_i x_i|$;
- L_1 -norm: if $X \in \mathbb{R}^N$, $|X| = \sum_{i=1}^N |x_i|$;
- e the normalized unit column vector: $1/N \times (1, \dots, 1)^T$;
- $G = \{G_0, G_1, G_2, \dots, G_n, \dots\}$ is a sequence of real vectors (in \mathbb{R}^N) such that $\sum_{n=0}^{\infty} |G_n| < \infty$;
- out_i is the number of non-zero entries of the i -th column of \mathbf{P} (counts outgoing links).

We assume in this paper that \mathbf{P} is non-negative for the sake of the simplicity. We could generalize some results below when $\rho(\tilde{\mathbf{P}}) \leq 1$, where $\tilde{\mathbf{P}}$ is the matrix where each component p_{ij} is replaced by $|p_{ij}|$.

2.1 Monotonicity

We say that \mathbf{P} is σ_v -decreasing if:

$$\forall X \in (\mathbb{R}^+)^N, \sigma_v(\mathbf{P}X) \leq \sigma_v(X).$$

We define $\mathbf{P}^\alpha = (1 - \alpha)\mathbf{I} + \alpha\mathbf{P}$.

Then, we have the following results:

PROPERTY 1. σ_v -decreasing property is stable by composition of operators (matrix product).

If \mathbf{P} is σ_v -decreasing, for all $\alpha \geq 0$, \mathbf{P}^α is σ_v -decreasing.

If \mathbf{P} is σ_v -decreasing, for all $(\alpha, \alpha') \in (\mathbb{R}^+)^2$ such that $\alpha \leq \alpha'$, $\sigma_v(\mathbf{P}^{\alpha'}X) \leq \sigma_v(\mathbf{P}^\alpha X)$.

PROOF. The first point is obvious. The other points are based on the linearity of σ_v .

2.2 Diffusion operators

We define the N diffusion operators associated to \mathbf{P} by:

$$\mathbf{P}_i = \mathbf{I} - \mathbf{J}_i + \mathbf{P}\mathbf{J}_i$$

PROPERTY 2. If \mathbf{P} is σ_v -decreasing, then the diffusion operators \mathbf{P}_i are σ_v -decreasing. Therefore, for $\alpha \geq 0$,

$$\mathbf{P}_i^\alpha = (\mathbf{P}_i)^\alpha = \mathbf{I} + \alpha(\mathbf{P} - \mathbf{I})\mathbf{J}_i$$

is σ_v -decreasing.

PROOF. $\sigma_v(\mathbf{P}_i X) = \sigma_v(X) + \sigma_v(\mathbf{P}\mathbf{J}_i X) - \sigma_v(\mathbf{J}_i X)$ and we have $\sigma_v(\mathbf{P}\mathbf{J}_i X) \leq \sigma_v(\mathbf{J}_i X)$, therefore $\sigma_v(\mathbf{P}_i X) \leq \sigma_v(X)$. The last point is the application of Property 1 to \mathbf{P}_i .

3. APPLICATION TO D-ITERATION

The D-iteration is defined by the couple $(\mathbf{P}, B) \in \mathbb{R}^{N \times N} \times \mathbb{R}^N$ and exploits two state vectors: H_n (history) and F_n (residual fluid) based on the following iterative equations:

$$\begin{aligned} F_0 &= B \\ F_n &= \mathbf{P}_{i_n} F_{n-1} \end{aligned} \quad (1)$$

and

$$\begin{aligned} H_0 &= 0 = (0, \dots, 0)^T \\ H_n &= H_{n-1} + \mathbf{J}_{i_n} F_{n-1}. \end{aligned} \quad (2)$$

The D-iteration consists in updating the joint iterations on (F_n, H_n) . Then, variant strategies may be applied depending on the choice of the sequence \mathcal{I} . To recall the dependence on \mathbf{P}, B and \mathcal{I} , we set: $H_n(\mathbf{P}, B, \mathcal{I}) = H_n$. When the limit is well defined we will set $H(\mathbf{P}, B, \mathcal{I}) = \lim_{n \rightarrow \infty} H_n(\mathbf{P}, B, \mathcal{I})$.

We will consider two cases: if $\rho(\mathbf{P}) < 1$ (ρ is the spectral radius of \mathbf{P}), then we will see that $H_n(\mathbf{P}, B, \mathcal{I})$ has a limit (denoted also H_∞) which is the unique solution of the equation (cf. Theorem 4):

$$X = \mathbf{P}X + B.$$

If $\rho(\mathbf{P}) = 1$, then we will only consider in this paper the D-iteration with $B = \mathbf{P}.e - e$ (or any other vector for which $\sigma_v(B) = 0$, with V a positive left eigenvector of \mathbf{P}).

3.1 General results

THEOREM 1 (FUNDAMENTAL DIFFUSION EQUATION). *We have:*

$$H_{n+1} + F_{n+1} = H_n + F_n + \mathbf{P}(H_{n+1} - H_n)$$

and

$$H_n + F_n = \mathbf{P}H_n + F_0. \quad (3)$$

PROOF. The first equation is straightforward from the equations (1) and (2). The second one can be obtained by induction.

The first equation means that what we have (sum of F and H) is what we had before plus what's diffused by the increment of H . The second equation means that what you have is the initial value plus what you received from diffusion.

THEOREM 2. *We have:*

$$H_n = (\mathbf{I} - \mathbf{J}_{i_n}(\mathbf{I} - \mathbf{P})) H_{n-1} + \mathbf{J}_{i_n} B.$$

PROOF. We can rewrite the equation as

$$H_n - H_{n-1} = \mathbf{J}_{i_n} (B - (\mathbf{I} - \mathbf{P})H_{n-1}).$$

Using (2), we only need to check that F_{n-1} is equal to $B - (\mathbf{I} - \mathbf{P})H_{n-1}$, which is exactly the equation (3).

3.2 Adding fluids

Consider the D-iteration $H_n(\mathbf{P}, B, \mathcal{I})$ on which we add G (we will denote this by $H_n(\mathbf{P}, B, \mathcal{I}, G)$): before each diffusion we add to F_n the vector G_n . This means that we modify F_n and H_n equations as follow:

$$F_n = \mathbf{P}_{i_n}(F_{n-1} + G_{n-1})$$

and

$$H_n = H_{n-1} + \mathbf{J}_{i_n}(F_{n-1} + G_{n-1}).$$

Then we have the following result:

THEOREM 3. *We have:*

$$H_{n+1} + F_{n+1} = H_n + F_n + G_n + \mathbf{P}(H_{n+1} - H_n)$$

and

$$H_n + F_n = \mathbf{P}H_n + F_0 + \sum_{i=0}^{n-1} G_i. \quad (4)$$

3.3 Convergence

We first show the convergence of $H_n(\mathbf{P}, B, \mathcal{I})$ when $\rho(\mathbf{P}) < 1$. This is a quite intuitive result and we only require that \mathcal{I} is a fair sequence.

DEFINITION 1. *A sequence \mathcal{I} is fair if the number of occurrences of each $i \in \Omega$ is unbounded.*

REMARK 1. *If we have $i \in \Omega$ such that $(F_n)_i$ is equal to zero after finite steps n_0 , we don't need the fairness condition for the position i (for the convergence).*

LEMMA 1. *If \mathbf{P} is irreducible and $\rho(\mathbf{P}) \leq 1$, then there exists a (strictly) positive vector V such that $|F_n|_v$ is non-increasing. As a consequence, F_n is convergent for any given sequence \mathcal{I} .*

PROOF. Set V the left positive eigenvector of \mathbf{P} for $\rho(\mathbf{P})$ (V is the left Perron vector, cf. [6], [4]). Set $j = i_{n+1}$ and $f = (F_n)_{i_{n+1}}$, then:

$$\begin{aligned} |F_{n+1}|_v &= \sum_{i \neq j} |(F_{n+1})_i v_i| + |(F_{n+1})_j v_j| \\ &= \sum_{i \neq j} |(F_n)_i + f p_{ij}| v_i + |f p_{jj}| v_j. \end{aligned}$$

Let's call Δ the set of nodes i such that $(F_n)_i$ has a sign opposed to f . Then,

$$\begin{aligned} |F_{n+1}|_v &= \sum_{i \neq j} (|(F_n)_i| + |f p_{ij}|) v_i + |f p_{jj}| v_j \\ &\quad + \sum_{i \in \Delta} (|(F_n)_i + f p_{ij}| - |(F_n)_i| - |f p_{ij}|) v_i \\ &= |F_n|_v + \sum_{i \in \Delta} (|(F_n)_i + f p_{ij}| - |(F_n)_i| - |f p_{ij}|) v_i \\ &\leq |F_n|_v. \end{aligned}$$

For the last inequality, we used $|x+y| \leq |x| + |y|$. Therefore, we have $|F_{n+1}|_v \leq |F_n|_v$. Therefore, F_n is convergent.

REMARK 2. *The above Lemma holds for any matrix \mathbf{P} , if there exists a positive vector V such that for all j , $\sum_i (|p_{ij}| \times v_i) \leq v_j$.*

LEMMA 2. *If $B = B_1 + B_2$, then for all n , we have*

$$H_n(\mathbf{P}, B, \mathcal{I}) = H_n(\mathbf{P}, B_1, \mathcal{I}) + H_n(\mathbf{P}, B_2, \mathcal{I}).$$

PROOF. The proof is straightforward using the linearity of H_n w.r.t. F_n .

3.3.1 Case $\rho(\mathbf{P}) < 1$

THEOREM 4. *If $\rho(\mathbf{P}) < 1$, for any fair sequence \mathcal{I} , $H_n(\mathbf{P}, B, \mathcal{I})$ is convergent to the unique vector X such that $X = \mathbf{P}X + B$.*

PROOF. Let's first assume that B is non-negative, so that we only manipulate non-negative quantities. By construction, H_n is non-decreasing per entry. From the equation (3), we have: $H_n = (\mathbf{I} - \mathbf{P})^{-1}(B - F_n)$. Hence, $H_n \leq (\mathbf{I} - \mathbf{P})^{-1}B$, therefore H_n is convergent and because \mathcal{I} is a fair sequence, necessarily, F_n tends to zero. Then, its limit satisfies the claimed equation. Now, if B has negative and positive terms, we can decompose B as $B^+ - B^-$ and apply the same argument for each component.

LEMMA 3. *If $\rho(\mathbf{P}) < 1$, then for all $(\alpha, \beta) \in \mathbb{R}^2$,*

$$\alpha H(\mathbf{P}, B, \mathcal{I}) + \beta H(\mathbf{P}, B', \mathcal{I}) = H(\mathbf{P}, \alpha B + \beta B', \mathcal{I}).$$

PROOF. We have:

$$\begin{aligned} H(\mathbf{P}, \alpha B + \beta B', \mathcal{I}) &= (\mathbf{I} - \mathbf{P})^{-1}(\alpha B + \beta B') \\ &= \alpha(\mathbf{I} - \mathbf{P})^{-1}B + \beta(\mathbf{I} - \mathbf{P})^{-1}B' \\ &= \alpha H(\mathbf{P}, B, \mathcal{I}) + \beta H(\mathbf{P}, B', \mathcal{I}). \end{aligned}$$

THEOREM 5 (SUPERPOSITION). *Let $S = \sum_{n=0}^{\infty} G_n$ with $\sum_{n=0}^{\infty} |G_n| < \infty$. If $\rho(\mathbf{P}) < 1$, then $H(\mathbf{P}, B, \mathcal{I}, G) = H(\mathbf{P}, B + S, \mathcal{I})$.*

PROOF. We have from the equation (4): $H_n(\mathbf{P}, B, \mathcal{I}, G) = (\mathbf{I} - \mathbf{P})^{-1}(B + \sum_{i=0}^{n-1} G_i - F_n)$. Then, one can easily check that the difference $|\sum_{i=0}^{n-1} G_i - F_n|$ tends to zero and the equality holds.

THEOREM 6 (MONOTONICITY). Let $S = \sum_{n=0}^{\infty} G_n$ with $\sum_{n=0}^{\infty} |G_n| < \infty$. Let $S' = \sum_{n=0}^{\infty} G'_n$ with $\sum_{n=0}^{\infty} |G'_n| < \infty$ such that $G_n \leq G'_n$. If $\rho(\mathbf{P}) < 1$, then for all n , $H_n(\mathbf{P}, B, \mathcal{I}, G) \leq H_n(\mathbf{P}, B, \mathcal{I}, G')$.

PROOF. The proof is straightforward by induction using the iterative equations of F_n and H_n ($F_n \leq F'_n$ and $H_n \leq H'_n$).

3.3.2 Case $\rho(\mathbf{P}) = 1$

In this case, we assume that the initial vector B satisfies $\sigma_v(B) = 0$, with V the left eigenvector of \mathbf{P} for eigenvalue 1.

Let's first consider the case when \mathbf{P} is irreducible. We denote by Ω_n^+ (resp. Ω_n^-) the subset of Ω such that $(F_n)_i \geq 0$ (resp. $(F_n)_i \leq 0$).

LEMMA 4. The diffusion operators preserve σ_v , which means that:

$$\sigma_v(\mathbf{P}_i X) = \sigma_v(X).$$

PROOF. $\sigma_v(\mathbf{P}_i X) = \sigma_v(X) + \sigma_v((\mathbf{P} - \mathbf{I})\mathbf{J}_i X)$. And we have $\sigma_v((\mathbf{P} - \mathbf{I})\mathbf{J}_i X) = (V^T \mathbf{P} - V^T)\mathbf{J}_i X = 0$.

LEMMA 5. For all n , $\sigma_v(F_n) = 0$.

PROOF. This is a direct consequence of Lemma 4 and $\sigma_v(F_0) = 0$ by assumption.

LEMMA 6. If at each diffusion step, $i_n \in \Omega_n^+$ (resp. $i_n \in \Omega_n^-$), then Ω_n^+ (resp. Ω_n^-) converges to Ω^+ (resp. Ω^-).

PROOF. It is clear that the subset Ω_n^+ is non-decreasing (we only add positive quantities to each node). It is bounded by Ω , hence convergent.

We denote by \mathcal{I}^+ (resp. \mathcal{I}^-) a fair sequence on Ω^+ (resp. Ω^-). $\Omega^+ = \cup_{n=1}^{\infty} \Omega_n^+$ (resp. $\Omega^- = \cup_{n=1}^{\infty} \Omega_n^-$), then, we have the following result.

THEOREM 7. If \mathbf{P} is irreducible and $\mathcal{I} = \mathcal{I}^+$, then $H_n(\mathbf{P}, \mathbf{P}e - e, \mathcal{I})$ is convergent to $X - e$ where X is the real right eigenvector of \mathbf{P} for the eigenvalue 1 with $\min_i x_i = 1/N$.

PROOF. If there exists $T < \infty$ such that $\cup_{n=1}^T \Omega_n^+ = \Omega$, then $\sigma_v(F_T^+) = \sigma_v(F_T^-) = 0$ and H_n converged in finite time. Otherwise, Ω^+ is strictly included in Ω . Let \mathbf{P}^+ be the restriction of \mathbf{P} on Ω^+ : then $\rho(\mathbf{P}^+) < 1$ (cf. [4]).

At the limit, H satisfies $(H + e) = \mathbf{P}(H + e)$ with $H + e$ a positive vector. There is at least one coordinate i on which the diffusion operator has been never applied (with positive fluid), therefore $\min_i (H)_i = 0$.

REMARK 3. If \mathbf{P} is not irreducible, the diffusion on \mathcal{I}^+ may not converge. The counter example is easy to be found.

In order to prove the convergence for the sequence \mathcal{I}^- without assuming the irreducibility of \mathbf{P} , we consider a bit more general diffusion iterations as follows:

$$F_n^\alpha = \mathbf{P}_{i_n}^{\alpha_n} F_{n-1}^\alpha \quad (5)$$

and

$$H_n^\alpha = H_{n-1}^\alpha + \alpha_n \mathbf{J}_{i_n} F_{n-1}^\alpha \quad (6)$$

where $\alpha_n \geq 0$. If for all n , $\alpha_n = 1$, we have the usual diffusion iteration.

THEOREM 8. (F_n^α, H_n^α) satisfies:

$$H_n^\alpha + F_n^\alpha = \mathbf{P} H_n^\alpha + B. \quad (7)$$

PROOF. The proof is the same as for the case $\alpha = 1$, by induction and using equations (5) and (6).

THEOREM 9. Assume we choose $F_0 \geq 0$ and $H_0 = 0$. Then, F_n^α and H_n^α are positive and $(H_n^\alpha)_i$ is an increasing function for all i .

If P is σ_v -decreasing, then $\sigma_v(F_n^\alpha)$ is a decreasing function.

PROOF. The proof is straightforward.

THEOREM 10 (PARTIAL DIFFUSION). If we build the two diffusion iterations (F_n^α, H_n^α) and $(F_n^{\alpha'}, H_n^{\alpha'})$ from the same initial vector F_0 ($H_0 = 0$) and for the same diffusion sequence \mathcal{I} , if for all n , $0 \leq \alpha_n \leq \alpha'_n \leq 1$, then we have:

- $\sigma_v(F_n^{\alpha'}) \leq \sigma_v(F_n^\alpha)$;
- $H_n^{\alpha'} \geq H_n^\alpha$ (for each vector entry);
- $H_n^{\alpha'} + F_n^{\alpha'} \geq H_n^\alpha + F_n^\alpha$.

PROOF. The first inequality is a direct consequence of Property 2 and Property 1. For the second and third inequalities, we prove by induction: we have obviously $H_0^{\alpha'} \geq H_0^\alpha$ and $H_0^{\alpha'} + F_0^{\alpha'} \geq H_0^\alpha + F_0^\alpha$. Assume we have, $H_n^{\alpha'} \geq H_n^\alpha$ and $H_n^{\alpha'} + F_n^{\alpha'} \geq H_n^\alpha + F_n^\alpha$. Then, from (6):

$$\begin{aligned} H_{n+1}^{\alpha'} &= H_n^{\alpha'} + \alpha'_{n+1} \mathbf{J}_{i_{n+1}} F_n^{\alpha'} \\ &\geq H_n^{\alpha'} + \alpha_{n+1} \mathbf{J}_{i_{n+1}} F_n^{\alpha'} \end{aligned}$$

and

$$H_{n+1}^\alpha = H_n^\alpha + \alpha_{n+1} \mathbf{J}_{i_{n+1}} F_n^\alpha$$

We first prove that:

$$H_n^{\alpha'} - H_n^\alpha \geq \alpha_{n+1} \mathbf{J}_{i_{n+1}} (F_n^\alpha - F_n^{\alpha'}).$$

For $i \neq i_{n+1}$, $(H_n^{\alpha'} - H_n^\alpha)_i \geq 0$ and $(\mathbf{J}_{i_{n+1}} (F_n^\alpha - F_n^{\alpha'}))_i = 0$. For $i = i_{n+1}$, we only need to handle the case $(F_n^{\alpha'} - F_n^\alpha)_{i_{n+1}} \geq 0$ and we use the relation: $H_n^{\alpha'} - H_n^\alpha \geq F_n^{\alpha'} - F_n^\alpha$. Hence, we have the inequality $H_{n+1}^{\alpha'} \geq H_{n+1}^\alpha$. Then, $H_{n+1}^{\alpha'} - H_{n+1}^\alpha \geq F_{n+1}^{\alpha'} - F_{n+1}^\alpha$ is straightforward using the equation (7) of Theorem 8.

REMARK 4. The power iteration $X_n = \mathbf{P} X_{n-1}$ can be described in the above scheme (F_n^α, H_n^α) with $X_n = F_{Nn}^\alpha$ taking $F_0 = X_0$ and if we apply the cyclic sequence $1, \dots, N$ ($i_n = n \bmod N$) where $\alpha_{kN+i} \leq 1$ is chosen such that we diffuse exactly $(\mathbf{P}^k X_0)_i$ (such a value exists and is less than 1 because after the diffusion of nodes $1, \dots, i$ the residual fluid on $(i+1)$ -th node can only be increased).

REMARK 5. Note that the above theorem is valid without any condition on $\rho(P)$.

THEOREM 11. Assume we have a strictly positive left-eigenvector $V > 0$ of \mathbf{P} . If we choose any fair sequence of nodes $\mathcal{I} = \mathcal{I}^-$ (we only diffuse negative fluids), then the diffusion applied on $(\mathbf{P}, \mathbf{P}e - e)$ converges to a unique H_∞ such that $H_\infty + e$ is the right-eigenvector for eigenvalue 1 for \mathbf{P} , such that the maximum value within each strongly connected component of spectral radius one is equal to $1/N$.

PROOF. The diffusion from $\mathbf{P}.e - e$ can be decomposed as the difference of two diffusion process $(F_n^{\alpha'}, H_n^{\alpha'})$ and (F_n^α, H_n^α) as follows: we start with $F_0 = e$. For the N first diffusions, we choose $i_n = n$ and

- for $\mathbf{P}_{i_n}^{\alpha_n}$, $\alpha_n = 0$;
- for $\mathbf{P}_{i_n}^{\alpha'_n}$, α'_n such that we diffuse exactly $1/N$ from all nodes (such a value exists and is less than 1 because after the diffusion of nodes $1, \dots, i$ the residual fluid on $(i+1)$ -th node can only be increased).

Then we have: $F_N^\alpha = e$, $H_N^\alpha = 0$ and $F_N^{\alpha'} = \mathbf{P}.e$, $H_N^{\alpha'} = e$. Then from the $(N+1)$ -th diffusion, we apply exactly the same sequence with $\alpha_n = \alpha'_n = 1$. For $n \geq N$, from Theorem 10 (and linearity of H , Lemma 2, w.r.t. $F_n = F_n^{\alpha'} - F_n^\alpha$), we have $H_n + e = H_n^{\alpha'} - H_n^\alpha \geq 0$ and we have $\sigma_v(F_n) = \sigma_v(F_n^{\alpha'} - F_n^\alpha) = 0$. If we only diffuse negative fluids, this means that H_n is a decreasing function (per entry). Since we have $0 \leq H_n + e \leq e$, H_n is convergent. We proved above (Lemma 1) that $|F_n|_v = \sum_i |v_i \times (F_n)_i|$ is a decreasing function. The convergence of H_n implies of course the convergence to zero of F_n on the coordinate $v_i > 0$ and we have $H + e = \mathbf{P}(H + e)$. Now we still have to prove that $H + e \neq 0$. In fact, there exists at least one entry such that $(H)_i = 0$: take $i \in \Omega_n^+$ (here we require that Ω_n^+ is only for entries strictly positive). If the iteration stops in finite time, the node in the last non empty Ω_n^+ satisfies $(H)_i = 0$. Otherwise, there exists at least one node such that it stays always strictly positive, for which we have $(H)_i = 0$. Hence, $\max_i (H + e)_i = 1/N$. The argument here applies for each strongly connected component (SCC) of spectral radius one (SCC(1)). For the uniqueness, we can decompose \mathbf{P} in three parts (cf. [4]):

- those who point to SCC(1)s and are not pointed to by any SCC(1)s: for them we have $(H)_i = 0$;
- those in SCC(1)s: for each SCC(1), H restricted on a SCC(1) is equal to the unique eigenvector of SCC(1) (Perron eigenvector) such that the maximum is equal to $1/N$;
- those who are pointed to by SCC(1)s and are not pointing to any SCC(1)s: this case corresponds to the previous section (diffusion on $\rho < 1$) for which the initial condition results from $\mathbf{P}e - e$ and the fluids received from SCC(1)s, on which we have uniqueness.

In order to justify clearly the above decomposition, we argue as follows: because of the existence of a strictly positive left-eigenvector of \mathbf{P} (for 1), we can decompose Ω as in Figure 1:

where the restriction of \mathbf{P} on the subsets IN and OUT has a spectral radius strictly less than 1.

The diffusion of nodes in IN is independent of the rest and its diffusion limit is $(H + e)_{i \in IN} = 0$ as shown in the following Lemma:

LEMMA 7. *If $\rho(\mathbf{P}) < 1$, and if we apply only negative diffusion from the initial condition $F_0 = \mathbf{P}.e - e$, then the limit is $H = -e$.*

PROOF. Since $\rho(\mathbf{P}) < 1$, the convergence of H is clear. We need to justify why diffusing only negative fluid, we do

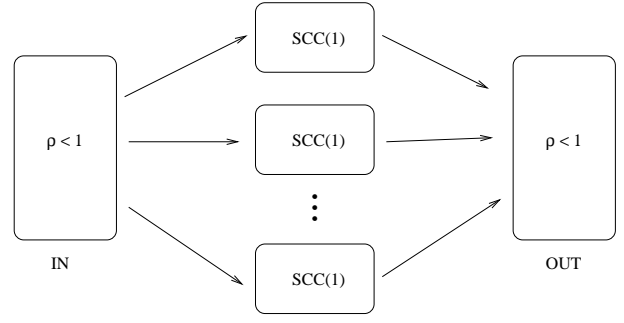


Figure 1: Decomposition.

not end up with some strictly positive coordinates. From Theorem 10, we have $H_n + e \geq 0$. Then from the equation (3), one can easily get: $\sigma_v(F_n) = (\rho(\mathbf{P}) - 1)\sigma_v(H_n + e)$, which means that F_n has always some negative components. At the limit, we can not have any negative fluids and $\sigma_v(F_\infty) = (\rho(\mathbf{P}) - 1)\sigma_v(H_\infty + e)$, which implies $H_\infty = -e$ and $F_\infty = 0$.

From the decomposition of Figure 1, once we have that $(H)_i + 1/N = 0$ on the nodes in IN, the diffusion of each SCC(1) is exactly the independent diffusion of \mathbf{P} restricted to each SCC(1) component, which converges to the unique eigenvector for 1. Finally, the nodes of OUT converges to a unique solution, since again we have $\rho(\text{OUT}) < 1$ and we can apply Theorem 5. This ends the proof of the theorem.

REMARK 6. *If V has some coordinates equal to zero, the above convergence works only on i such that $v_i > 0$ and we can have $(F_\infty)_i > 0$ on i such that $v_i = 0$. Let $\Omega^{+\infty}$ be the set coordinates such that $(F_\infty)_i > 0$. Then the spectral radius of P restricted to $\Omega^{+\infty}$ may be one, in which case, there is no convergence.*

REMARK 7. *If we mix the diffusion of positive and negative fluids, there is no guarantee that the D-iteration algorithm converges. For instance, with a snake-configuration counter example, we may oscillate (take for instance, \mathbf{P} associated to a five nodes graph: $1 \rightarrow 2$ and $1 \rightarrow 3$; $2 \rightarrow 4$; $3 \rightarrow 5$; $4 \rightarrow 1$; $5 \rightarrow 1$).*

THEOREM 12 (MONOTONICITY 2). *Let $S = \sum_{n=0}^{\infty} G_n$ with $\sum_{n=0}^{\infty} |G_n| < \infty$. Let $S' = \sum_{n=0}^{\infty} G'_n$ with $\sum_{n=0}^{\infty} |G'_n| < \infty$. If for all n , $\sum_{i=0}^n G_i \leq \sum_{i=0}^n G'_i$, then for all n , $H_n(\mathbf{P}, B, \mathcal{I}, G) \leq H_n(\mathbf{P}, B, \mathcal{I}, G')$.*

PROOF. This is a particular case of the partial diffusion result (Theorem 10): $H_n(\mathbf{P}, B, \mathcal{I}, G)$ can be seen as $H_n(\mathbf{P}, B + G', \mathcal{I})$ where at step n , $\sum_{i=n+1}^{\infty} G'_i + \sum_{i=0}^n (G'_i - G_i)$ has been blocked. $H_n(\mathbf{P}, B, \mathcal{I}, G')$ can be seen as $H_n(\mathbf{P}, B + G', \mathcal{I})$ where at step n , $\sum_{i=n+1}^{\infty} G'_i$ has been blocked.

4. CONVERGENCE SPEED

For the sake of the simplicity, we only considered here the case of sub-stochastic matrices of the form: $\mathbf{P} = d.\mathbf{P}_s$ where $0 < d < 1$ and \mathbf{P}_s is a transition matrix.

We consider the following choice of sequence:

- CYC: $i_n = n \mod N$;
- MAX: $i_n = \arg \max_i (F_n)_i$;
- COST: $i_n = \arg \max_i ((F_n)_i / \text{out}_i)$.

Let's first consider the theoretical cost of the iteration as the number of times diffusions are applied.

THEOREM 13. *The convergence speed of CYC and MAX is at least $d^{\lfloor l/N \rfloor}$, where l is the number time diffusions are applied.*

PROOF. CYC: from F_n , after N diffusions, we diffused at least $|F_n|$, therefore, $|F_{n+N}| \leq d|F_n|$.

MAX: if we order $(F_n)_i$: $(F_n)_1 \geq (F_n)_2 \geq \dots (F_n)_N$, we start by diffusing $(F_n)_1$, and at the k -th diffusion we take a fluid at least equal to the k -th term $(F_n)_k$. As for CYC, we diffused at least $|F_n|$ after N diffusions.

Now let's consider the theoretical cost of the iteration as the number of times a non-zero element of P is used.

THEOREM 14. *The convergence speed of COST is at least $d^{\lfloor L/l \rfloor}$, where $L = \sum_i \text{out}_i$ and l is the number of times a non-zero element of \mathbf{P} is used.*

PROOF. We just need to count as above the amount of fluid we move while using L links of P . At step n , assume we order $(F_n)_i/\text{out}_i$: $(F_n)_1/\text{out}_1 \geq (F_n)_2/\text{out}_2 \geq \dots \geq (F_n)_N/\text{out}_N$. At $n+1$, we diffuse exactly $(F_n)_1$ which costs out_1 . Assume we have diffused D_{k-1} with L_{k-1} operations. Then, At step $n+k$, we diffuse d_k with cost l_k such that $d_k/l_k \geq (F_n)_k/\text{out}_k$. Hence, we have $D_k/L_k \geq (\sum_{i=1}^k (F_n)_i)/(\sum_{i=1}^k \text{out}_i) \geq |F_n|/L$. If we stop counting when L_k just exceeds L (say k_0), then, $D_{k_0} \geq |F_n|/L \times L_{k_0} \geq |F_n|$.

REMARK 8. *An interesting heuristic bound is $d/(2-d)$ assuming both the graph and the sequence are random.*

5. CONVERGENCE OF DISTRIBUTED ALGORITHM

5.1 Case $\rho < 1$

Here we first consider the case $\rho(\mathbf{P}) < 1$. We detail the proof of the convergence for $K = 2$ when the computation is done on two virtual processors, that we call PID1 and PID2. We assume a fixed partition of $\Omega = \Omega_1 \cup \Omega_2$ and the corresponding decomposition of \mathbf{P} as:

$$P = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}$$

The diffusion state variables of two PIDs are denoted by: (F^1, H^1) , (F^2, H^2) . For a single PID sequential system, we note $H = ([H]^1, [H]^2)$.

To prove the convergence of the distributed architecture to the proper eigenvector, we consider the following decomposition: we assume that PID_1 computes the D-iteration $H(\mathbf{P}_{11}, B_1, \mathcal{I}^1)$ with additional inputs from PID_2 at times T_1, \dots, T_n, \dots . During $\tau_n = T_n - T_{n-1}$ (we set $T_0 = 0$), PID_1 applies the diffusion sequence \mathcal{I}_n^1 . We assume without any loss of generality that PID_2 computes the D-iteration based on the sequence \mathcal{I}^2 (\mathcal{I}_n^2 during $T'_n - T'_{n-1}$) and sends at T'_n ($n \geq 1$) the quantity

$$\mathbf{P}_{12}(H_{T'_n}^2 - H_{T'_{n-1}}^2) \stackrel{\text{def}}{=} \Delta H_n^2.$$

Let's first assume that this quantity defines the additional input at T_n by PID_1 (one way). The information transmission delay is then defined by $T_n - T'_n$. We will see that we don't need that the information from PID_2 arrives in the order for the proper convergence. The only requirement is that $T_n - T'_n$ and $T'_n - T'_{n-1}$ are bounded.

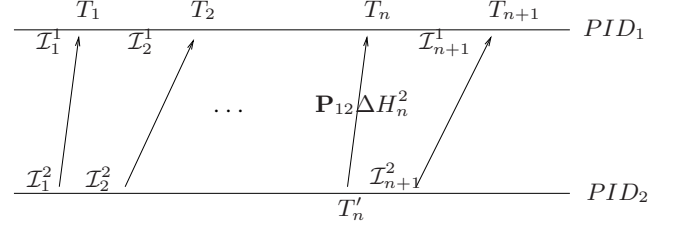


Figure 2: Data exchange: from PID2 to PID1.

If n is such that $T_k \leq n < T_{k+1}$, we have (upper bound):

$$\begin{aligned} H_n^1 &= H_{T_1}(P_{11}, B_1, \mathcal{I}_1^1) \\ &\quad + H_{\tau_2}(P_{11}, F_{T_1}^1 + \Delta H_1^2, \mathcal{I}_2^1) \\ &\quad \dots \\ &\quad + H_{\tau_k}(P_{11}, F_{T_{k-1}}^1 + \Delta H_{k-1}^2, \mathcal{I}_k^1) \\ &\quad + H_{n-T_k}(P_{11}, F_{T_k}^1 + \Delta H_k^2, \mathcal{I}_{k+1}^1) \\ &= (\mathbf{I} - \mathbf{P}_{11})^{-1}(B_1 + \mathbf{P}_{12}H_{T_k}^2 - F_n^1). \end{aligned}$$

Hence,

$$H_n^1 \leq (\mathbf{I} - \mathbf{P}_{11})^{-1}(B_1 + \mathbf{P}_{12}H_n^2). \quad (8)$$

In the above, we used the equality

$$H_{\tau_i}(\mathbf{P}_{11}, F_{T_{i-1}}^1 + \Delta H_{i-1}^2, \mathcal{I}) = (\mathbf{I} - \mathbf{P}_{11})^{-1}(F_{T_{i-1}}^1 + \Delta H_{i-1}^2 - F_{T_i}^1).$$

The above result shows that only the cumulated fluids matter for PID1.

Let's now use the notation of the G (fluid addition): if $n = T_i$, we set $G_n^2 = \Delta H_i^2$, otherwise $G_n^2 = 0$. We assume $S = \sum_{i=1}^{\infty} G_i^2 = \mathbf{P}_{12}H_{\infty}^2 < \infty$. Then $H_n^1 = H_n(\mathbf{P}_{11}, B_1, \mathcal{I}^1, G^2)$ (we define G_n^1 in the same way). From Theorem 5, we have $H_{\infty}^1 = H(\mathbf{P}_{11}, B_1 + G, \mathcal{I}^1)$.

Now consider the global system. Let's denote by (T'_n, T_n) the sending time and the receiving time of the n -th fluid transmission exchange between PIDs (in both directions). Because the different PIDs may have different computation speed, we may have that at T_n the number of iterations done by the PID that receives the fluid is less than the number of iterations done by the sender at time T'_n . For the sake of the simplicity, we still denote by F_{T_n} (same for H and for T'_n) the value of the F at time T_n .

LEMMA 8 (IMPACT OF THE DELAY). *Consider two distributed computation systems $S(1), S(2)$ that have exactly the same T'_n and that differs only by T_n such that for all n , $T_n(1) \leq T_n(2)$. Then, we have: $H_n^1(1) \geq H_n^1(2)$ and $H_n^2(1) \geq H_n^2(2)$.*

PROOF. This is a direct consequence of Theorem 12.

LEMMA 9. *At each step, he have $H_n^1 \leq [H]_n^1$ and $H_n^2 \leq [H]_n^2$. Therefore, the distributed computation scheme converges.*

PROOF. From Lemma 8, we first have the monotonicity when reducing the delay up to zero delay $T_n = T'_n$. Then, again thanks to Theorem 12, one can show that adding more fluid exchanges, we can only increase the history vector. The system with zero delay and exchange at each step corresponds to the sequential system H . Therefore, we have $H_n^1 \leq [H]_n^1 \leq [H]_\infty^1$ and $H_n^2 \leq [H]_n^2 \leq [H]_\infty^2$, and H_n^1 and H_n^2 are convergent.

THEOREM 15. *If each PID applies a fair sequence for diffusions and if for all n , $|T_n - T'_n| \leq T$ and $|T'_n - T'_{n-1}| \leq T$, then the distributed computation converges to the sequential single PID computation H .*

PROOF. From the equation (8), we still have: $H_n^1 \leq (\mathbf{I} - \mathbf{P}_{11})^{-1}(\mathbf{B}_1 + \mathbf{P}_{12}H_\infty^2)$ and $H_n^2 \leq (\mathbf{I} - \mathbf{P}_{22})^{-1}(\mathbf{B}_2 + \mathbf{P}_{21}H_\infty^1)$.

If for all n , $|T_n - T'_n| < T$ and $|T'_n - T'_{n-1}| < T$ we have (lower bound):

$$H_{n+2T}^1 \geq (\mathbf{I} - \mathbf{P}_{11})^{-1}(\mathbf{B}_1 + \mathbf{P}_{12}H_n^2 - F_{n+2T}^1)$$

and

$$H_{n+2T}^2 \geq (\mathbf{I} - \mathbf{P}_{22})^{-1}(\mathbf{B}_2 + \mathbf{P}_{21}H_n^1 - F_{n+2T}^2)$$

We already know that both converges (non-decreasing and upper bounded by $[H]_\infty^1$ and $[H]_\infty^2$), therefore F_{n+2T}^1 and F_{n+2T}^2 goes to zero and at the limit we have:

$$H_\infty^1 = (\mathbf{I} - \mathbf{P}_{11})^{-1}(\mathbf{B}_1 + \mathbf{P}_{12}H_\infty^2)$$

and

$$H_\infty^2 = (\mathbf{I} - \mathbf{P}_{22})^{-1}(\mathbf{B}_2 + \mathbf{P}_{21}H_\infty^1).$$

Therefore, we have $H_\infty^1 = [H]_\infty^1$ and $H_\infty^2 = [H]_\infty^2$ (because H is the unique solution satisfying the above two equations).

Now, the generalization of this proof to any $K > 1$ follows exactly the same ideas.

5.2 Case $\rho = 1$

Here, we assume that there exists V a strictly positive left eigenvector of \mathbf{P} and that the initial vector B satisfies $\sigma_v(B) = 0$. Let's consider the limit of the diffusion on Ω^- .

We use the same notation than in the previous section. Since we only apply the diffusion on nodes having negative fluids, the history vector H is for each entry a negative decreasing function.

In the previous case with $\rho < 1$, we could assume that \mathcal{I}^1 and \mathcal{I}^2 are predefined and independent of the fluid exchanges. The first difficulty here is that a priori we can no more do such an assumption, since the sign of $(F)_i$ depends on the past fluid exchanges.

To overcome this difficulty and apply the previous results, we do the following *trick* which consists in the decomposition of the diffusion in two steps: selection of a node and diffusion test. We assume given two fair sequences \mathcal{I}^1 and \mathcal{I}^2 on Ω_1 and Ω_2 (fair means here that every coordinate will be candidate for diffusion an infinite number of times) and when a node is selected, we only diffuse if its sign is negative.

Then, we have the following result:

THEOREM 16 (MONOTONICITY EXTENDED). *Let G_n and G'_n two sequences of non-positive vectors, such that for all n , $\sum_{i=1}^n G_i \leq \sum_{i=1}^n G'_i$. Let \mathcal{I} be a fair sequence. If we only diffuse the negative fluids, then for all n , $H_n(\mathbf{P}, B, \mathcal{I}, G) \leq H_n(\mathbf{P}, B, \mathcal{I}, G')$.*

PROOF. The argument is exactly the one used in Theorem 12.

Then, the results of Lemma 8 and Lemma 9 still hold with inversed inequality.

THEOREM 17. *If each PID applies a fair sequence for diffusions and if for all n , $|T_n - T'_n| \leq T$ and $|T'_n - T'_{n-1}| \leq T$, then the distributed computation on Ω^- converges to the sequential single PID computation H obtained by the diffusion of negative fluid.*

PROOF. We have the lower bounds: $[H]_\infty^1 \leq [H]_n^1 \leq H_n^1$ and $[H]_\infty^2 \leq [H]_n^2 \leq H_n^2$, therefore H_n^1 and H_n^2 are convergent.

We know that the limit of the distributed computation satisfies:

$$(\mathbf{I} - \mathbf{P}_{11})H_\infty^1 = \mathbf{B}_1 + \mathbf{P}_{12}H_\infty^2 \quad (9)$$

$$(\mathbf{I} - \mathbf{P}_{22})H_\infty^2 = \mathbf{B}_2 + \mathbf{P}_{21}H_\infty^1. \quad (10)$$

Let $Z = ((H_\infty^1)^T, (H_\infty^2)^T)^T - H$. If $\rho(\mathbf{P}_{11}) < 1$ and $\rho(\mathbf{P}_{22}) < 1$, then we have the uniqueness of the solution and $Z = 0$. If $\rho(\mathbf{P}_{11}) = 1$ and $\rho(\mathbf{P}_{22}) = 1$, then since $\rho(\mathbf{P}) = 1$ and since we assumed the existence of strictly positive left-eigenvector for 1, we can not have any positive terms linking Ω_1 and Ω_2 (cf. [4]), hence we have two independent systems for which the uniqueness was proved (Theorem 11). Finally, if $\rho(\mathbf{P}_{11}) = 1$ and $\rho(\mathbf{P}_{22}) < 1$, then we must have $\mathbf{P}_{12} = 0$, and for the same reason, we still have $Z = 0$.

6. EXPERIMENTATION

6.1 Uniform graph with $N = 128$

We first report the experiment results by reproducing the hypothetical parallel computer simulations as defined in [15] (Table 1). We reproduced a random (symmetrical) graph of 128 nodes containing 1652 links, 23 of which are self-loops. As in [15], we assumed that a processor spends 0 cycles to read and write its local memory, T_r and T_w cycles, respectively, to read from and write into the shared memory, T_m cycles for one multiplication and T_a cycles for one addition (below, we took $T_r = 4$, $T_w = 2$, $T_m = T_a = 1$). The degrees of the nodes vary from 14 to 38, with a mean of 25.6 and standard deviation of 4.9. In Table 1, we show the results of:

- sPI-R: synchronized power iteration per row (i.e. Jacobi iteration);
- aPI-R: asynchronized power iteration per row (method evaluated in [15]);
- sPI-C: synchronized power iteration per column;
- sPI-Cr: synchronized power iteration per column assuming identical weight p_{ij} for each column j (as in PageRank equation);
- DI+Cost: asynchronized diffusion applying COST from the initial vector $\mathbf{P} \cdot e - e$ (cost of the computation of $\mathbf{P} \cdot e - e$ is included).

We see that the results of sPI-R and aPI-R are very close to those observed in Table 1 of [15]. With sPI-C, we see a

K	sPI-R	aPI-R	sPI-C	sPI-Cr	DI+COST
1	65620	45934	65620	34090	18183
2	66020	45472	37500	21470	20047
4	41620	28504	23180	15030	17347
8	24000	16590	16000	11770	13711
16	13060	9142	11940	9780	9952
32	7880	5056	9980	8720	6469
64	4120	3056	8210	7550	4264
128	2260	2256	7215	6845	3128

Table 1: Uniform graph $N = 128$.

gain for $K = 2, 4, 8$: that's because the fluid exchange to other PIDs aggregate the results on the diffusions of N/K nodes to $N/K * (K - 1)$ nodes. However, there is a higher penalty when K becomes close to N . This can be explained considering the limit case of $K = N$: whereas one coordinate updates on a row having n_r non-zero values requires about $2 \times n_r$ additions and multiplications followed by one writing cycle, the diffusion would require $2 \times n_c$ additions and multiplications (for a column having n_c non-zero values) followed by n_c updates (read, addition and write) of shared memory. The result for sPI-Cr shows the gain when exploiting the homogeneity of the column weight (one multiplication required per diffusion), which can be done only when column based operations are used (as for the diffusion approach).

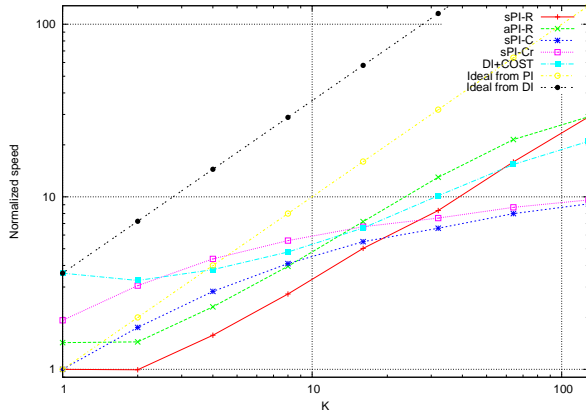


Figure 3: Normalized speeds.

Figure 3 shows the results of Table 1 in terms of the convergence speeds normalized by the value of sPI-R for $K = 1$. Ideal-PI and Ideal-DI are the ideal curve ($y = x$) starting from $K = 1$ of PI and DI.

6.2 Web graph

In this section, we used the web graph imported from the dataset `uk-2007-05 @1000000` (available on [1]) which has 41,247,159 links on 10^6 nodes. The aim is to compare the theoretical computation cost for a large sparse matrix such as the one associated to a web graph, the case for which the diffusion approach was initially designed. The $N = 1000$ case is obtained considering the first 1000 nodes of the above graph. Here the convergence is on the PageRank equation (eigenvector with damping factor of 0.85). The results are shown on Figure 4 for $N = 1000$ and Figure 5 for $N = 10^6$.

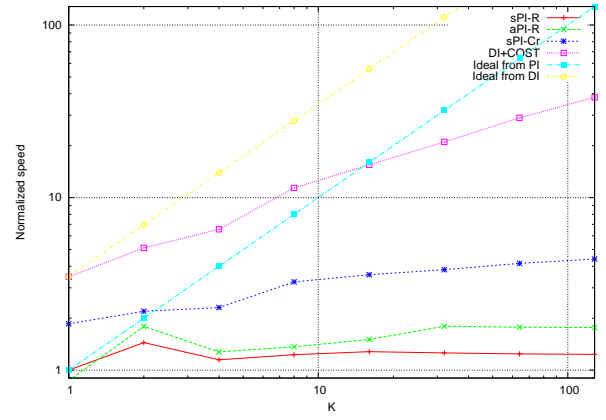


Figure 4: Normalized speeds: $N = 1000$.

In Figure 4, the benefit of the column based diffusion approach is much more significant and even more significant for $N = 10^6$ (Figure 5). In Figure 5, we added the performance of a dynamical partition approach (cf. [10] for more details: the dynamical partition used here consists roughly in observing the convergence speed in logscale of each PID based on its remaining fluid quantity and transferring 10% of nodes that is managed by the slowest PID to the fastest PID when the difference is higher than 50%) in order to check/validate the property claimed in [10]: we can see that applying the cost assumption/model of 4, a simple dynamical partition strategy on the diffusion method leads to a performance that is close to the optimal efficiency (close to the ideal curve). Note that the cost of the partition updates/modifications is partially included here (unity cost per node exchanged on both sides of PIDs).

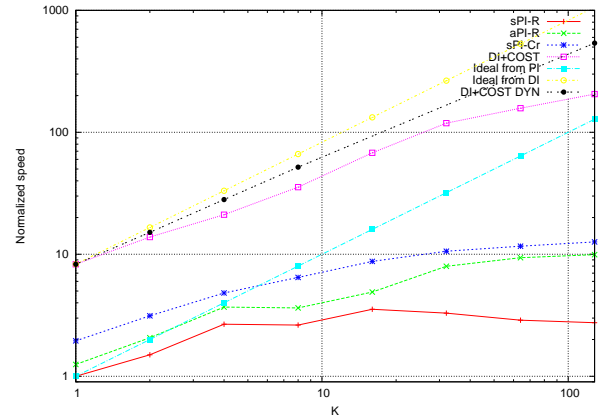


Figure 5: Normalized speeds: $N = 10^6$.

Acknowledgments

The authors are very grateful to Philippe Jacquet for his very valuable comments and suggestions.

7. CONCLUSION

In this paper, we addressed the formal proof of convergence of different D-iteration schemes, including the case

of the distributed computation and the upper bounds of the convergence speeds. We used the theoretical formal approach of [15] to evaluate the theoretical computation cost and showed the significant gain of our diffusion approach for the computation of the eigenvector of large sparse matrices.

8. REFERENCES

- [1] <http://law.dsi.unimi.it/datasets.php>.
- [2] S. Abiteboul, M. Preda, and G. Cobena. Adaptive on-line page importance computation. *WWW2003*, pages 280–290, 2003.
- [3] W. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.
- [4] A. Berman and A. Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences: Abraham Berman, Robert J. Plemmons*. Number pt. 11 in Classics in Applied Mathematics Series. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1994.
- [5] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, Jan. 2008.
- [6] F. Gantmacher. *The theory of matrices. 2*. Chelsea Publishing Series. AMS Chelsea Pub, 2000.
- [7] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.
- [8] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, 1997.
- [9] E. Hestenes, Magnus R.; Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6).
- [10] D. Hong. D-iteration: Evaluation of a dynamic partition strategy. *Proc. of AHPCN 2012*, June 2012.
- [11] D. Hong. D-iteration: Evaluation of the asynchronous distributed computation. <http://arxiv.org/abs/1202.6168>, February 2012.
- [12] D. Hong. D-iteration method or how to improve gauss-seidel method. *arXiv*, <http://arxiv.org/abs/1202.1163>, February 2012.
- [13] D. Hong. Optimized on-line computation of pagerank algorithm. <http://arxiv.org/abs/1202.6158>, 2012.
- [14] J. Krüger and R. Westermann. Linear algebra operators for gpu implementation of numerical algorithms. *ACM Trans. Graph.*, 22(3):908–916, July 2003.
- [15] B. Lubachevsky and D. Mitra. A chaotic asynchronous algorithm for computing the fixed point of a nonnegative matrix of unit spectral radius. *J. ACM*, 33(1):130–150, Jan. 1986.
- [16] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.
- [17] G. Stewart. *Matrix Algorithms, Volume 1 to 5*. SIAM, Philadelphia, 2002.
- [18] R. Varga. *Matrix Iterative Analysis*. Springer Series in Computational Mathematics. Springer, 2009.